

Aplicació per a mòbils Android per a poder cercar text en imatges

Eloi Vila Tudela

Resum—Les imatges s'usen cada vegada més en àmbits molt diversos. Aquestes imatges solen contenir colors, objectes i/o textos i les bases de dades poden contenir infinitat d'imatges. Existint tantes imatges apareix la necessitat de classificar-les, d'ordenar-les i de poder-ne buscar alguna en concret i trobar-la amb relativa facilitat (que una persona hagi de repassar totes les imatges d'una base de dades pot ser una tasca molt poc productiva i ineficient). El resultat d'aquest treball és la classificació d'imatges segons les lletres o text que contenen i la seva posterior recuperació. S'ha desenvolupat una aplicació per a dispositius mòbils amb sistema *Android* que envia el text que l'usuari vol buscar al servidor on estan allotjades les imatges. Aquest servidor busca entre les imatges que disposa les cinc que contenen un text i que aquest s'assembla més al text rebut. Aquestes imatges són retornades a l'aplicació i mostrades a l'usuari. Un exemple d'ús podrien ser les imatges del sistema *Street View* de Google.

Paraules clau—Android, aplicació, imatge, text.

Abstract—The images are used more and more in many different scopes. These images usually contain colors, objects and/or texts and the databases can contain countless images. With this growing number of images it appears the necessity to classify them, order them and be able to look for one in particular and find it with relative ease (if a person has to go through all the images of a database can become an unproductive and inefficient task). The outcome of this project is the classification of images according the letters or text containing and the subsequent recovery. We have developed an application for mobile devices with *Android* system which sends the text the user wants to search to the server where the images are hosted. This server searches through the five available images containing text and this text looks more like the received text. These images are returned to the application and displayed to the user. An example of use could be the images of the Google *Street View* system.

Index Terms—Android, application, image, text.



1 INTRODUCCIÓ

AQUEST document és l'article que resumeix la realització del treball de final de grau "Aplicació per a mòbils Android per a poder cercar text en imatges". En ell es recullen els objectius que es pretenia assolir i la motivació que els impulsava. Es tracten punts importants del projecte com la metodologia, la planificació i una explicació dels diferents punts claus per al desenvolupament del mateix (en què es mostren captures de pantalla de la interfície desenvolupada). Finalment, es recullen i discuteixen els resultats aconseguits i es tanca l'article amb les conclusions obtingudes en la realització.

1.1 Motivació

Amb l'augment del nombre d'imatges que s'utilitzen en cada vegada més àmbits apareix una necessitat, i és la de poder buscar i recuperar aquestes imatges de forma ràpida i senzilla. Si estem treballant en conjunts de poques imatges no tenim gaire problema, però si aquests petits conjunts es transformen en grans bases de dades la cosa canvia. La principal motivació d'aquest treball és solucionar aquest problema, és a dir, proveir una solució per a poder filtrar grans conjunts d'imatges i evitar perdre molt de temps en la cerca. Es podrien utilitzar filtres molt diversos, ja sigui si contenen un color determinat o no, si contenen molts colors o estan en blanc i negre, etc. El filtre utilitzat en aquest treball és el text, es buscaran imatges que continguin un text igual o molt semblant al text introduït per l'usuari. Abans de començar el projecte s'ha volgut analitzar qui hi podria estar interessat:

Possibles usuaris potencials

- Homes i dones d'entre 15 i 55 anys, que tenen imatges guardades i volen classificar-les o buscar-ne alguna de concreta (en el seu temps d'oci, no professionalment).

-
- E-mail de contacte: Eloi.VilaT@e-campus.uab.cat
 - Menció realitzada: Enginyeria del Software i Tecnologies de la Informació.
 - Treball tutoritzat per: Ernest Valveny Llobet (Departament de Ciències de la Computació)
 - Curs 2013/14

- Persones que treballin amb grans bases de dades d'imatges i necessitin recuperar-ne alguna amb un text determinat.

Altres parts interessades

- Empreses que treballin amb grans bases de dades d'imatges que continguin textos.
- Empreses dedicades al tractament d'imatges.
- Empreses dedicades a la investigació o al desenvolupament d'aplicacions semblants.

Elecció plataforma

En la primera reunió sobre el projecte, abans de començar a parlar sobre temes de desenvolupament, convenia decidir si l'aplicació seria una aplicació per a dispositius mòbils o si seria una aplicació web. A priori podia semblar que una aplicació web podria aportar més facilitats però amb una aplicació mòbil preveiem que seria més còmode i usable per part de l'usuari. Les dues opcions eren vàlides però per falta de temps era impossible abordar les dues. Així doncs, com que l'estudiant tenia més motivació per desenvolupar per a dispositius mòbils, aquesta plataforma va ser l'escollida. Existeixen diferents sistemes mòbils, però aquí no hi va haver discussió ja que es va considerar que *Android* oferia més facilitats al desenvolupador i disposava de més usuaris que els seus competidors. Una vegada la tria estava feta es van incorporar els canvis al títol i es va encaminar el projecte seguint aquestes decisions.

1.2 Producte i documentació del projecte

Descripció del producte

El producte obtingut del projecte és el conjunt format per una aplicació per a dispositius mòbils *Android* i un servidor amb les funcions per a reconèixer on es troba el text en una imatge. Aquesta aplicació permet a l'usuari buscar si alguna de les imatges que té guardades a la base de dades conté el text que ell haurà introduït prèviament. El sistema interactua amb el servidor i aquest retorna les imatges que contenen el text més semblant amb un possible marge d'error (el reconeixement de text no és exacte, ja que les imatges podrien tenir alguna part borrosa, diferents estils de lletres, mides, obstacles en la imatge que dificultin la visió del text sencer, etc.).

Accions de documentació

Durant la realització d'aquest treball s'han generat els següents documents: l'informe inicial (per a definir els objectius i l'abast del projecte així com els recursos disponibles, les tasques principals i una planificació temporal per a dur-les a terme), dos informes de progrés (realitzats per a veure els progressos aconseguits passats un i dos mesos de l'inici del projecte) i aquest article (que recull punts importants definits en els informes i una explicació clara sobre què consisteix el treball).

Accions de comunicació

- Reunions amb el tutor del treball.
- Reunions amb els estudiants que han col·laborat en el projecte.
- Informes periòdics de seguiment del projecte.

1.3 Objectius del treball

Aquest projecte té tres objectius principals. El primer és disposar d'un servidor per allotjar les imatges. Aquest servidor ha de contenir les funcions per analitzar on es troba el text en una imatge i les funcions per a poder interpretar aquesta informació i decidir de quin text es tracta (aquestes funcions no s'implementaran en aquest treball). El segon objectiu és disposar d'una aplicació mòbil per a l'usuari. L'usuari ha de poder entrar un text, enviar-lo al servidor i veure les imatges que aquest l'hi ha retornat. El tercer objectiu és que hi hagi una comunicació eficient i fiable entre el servidor i l'aplicació.

Exposats els principals objectius, alguns requisits importants per al projecte són:

1. Disposar d'una interfície clara i entenedora per a que l'usuari pugui introduir el text a buscar.
2. Realitzar tots els càlculs possibles al servidor per a no sobrecarregar el dispositiu.
3. L'aplicació ha de funcionar en sistemes mòbils i tauletes *Android*.

No objectius

Donada la complexitat dels mateixos i el poc temps per a realitzar el projecte, convé destacar dos aspectes importants que no es desenvoluparan en aquest projecte.

1. No s'implementaran les funcions per a poder reconèixer si una imatge conté un text i on es troba (proporcionen les coordenades del rectangle que conté el text).
2. No s'implementaran les funcions que es relacionen amb les anteriors i ens donen informació sobre el text contingut en el rectangle.

Com que aquestes funcions són vitals per al projecte i s'han d'utilitzar han estat implementades per dos estudiants de doctorat com a col·laboració en el projecte.

1.4 Estat de l'art

En els últims anys la utilització dels aparells mòbils ha anat en augment. Una de les raons és l'evolució que han sofert aquests dispositius, que ja no es basen només en la possibilitat de realitzar trucades des de qualsevol lloc. Inclouen noves prestacions que els fan més atractius i més útils per als usuaris, prestacions que oferien altres aparells i ara estan centralitzades en un de sol (possibilitat de tirar fotografies, gravar vídeos, connectar-se a Internet, obtenir coordenades i mapes, etc.). Tant atractives són aquestes prestacions que han anat desplaçant la principal, realitzar trucades. Aquest desplaçament és fa evident amb l'aparició de les tauletes, "mòbils" que ofereixen totes aquestes prestacions (amb algunes millores, com per exemple una pantalla més gran o més capacitat) sense oferir la possibilitat de trucar. Això ha obligat al concepte de mòbil a evolucionar per incloure aquests nous models, parlarem de dispositius mòbils.

Totes aquestes prestacions són ofertes a l'usuari a través d'aplicacions, l'aparició de les quals també està en augment i se'n poden trobar de molt diverses. L'aplicació realitzada en aquest projecte intenta resoldre el problema de cerca d'imatges, què fer quan tenim moltes imatges i en volem buscar una en concret. Els filtres de cerca podrien ser molt diversos, ja sigui per mida, colors o qualitat de la imatge, però el filtre escollit en aquest cas és el text,

és a dir, si les imatges contenen algun tipus de text en concret o algun de semblant.

Com els mòbils i les aplicacions, el concepte d'imatge també ha evolucionat i cada vegada apareixen més estudis relacionats amb el tema. En aquest treball hem aprofitat els mètodes descrits en [3] i [4], articles en què han treballat els estudiants que han col·laborat en el projecte realitzant les funcions necessàries per al tractament d'imatges.

2 METODOLOGIA

Com a metodologia de gestió de software s'han utilitzat algunes característiques de la metodologia coneguda com a Programació Extrema o Extreme Programming. Xp és un procés àgil de desenvolupament que inclou diverses característiques interessants per a aquest projecte, com són el desenvolupament iteratiu i incremental, proves contínues (s'aconsella escriure el codi de prova abans que la funció en si) i la correcció de tots els errors abans d'afegir una funcionalitat nova. No s'ha pogut utilitzar la totalitat de la metodologia perquè es recomana que els equips estiguin formats per entre 2 i 12 membres. Aquest inconvenient fa que característiques com la programació per parelles, per exemple, no s'hagi pogut aplicar perquè el projecte ha estat desenvolupat per un sol estudiant.

Les funcions de reconeixement de text estan escrites en llenguatge C++, ja que és el llenguatge orientat a objectes que disposa de més llibreries i facilitats per al tipus de projecte a desenvolupar. Projectes anteriors relacionats amb el tema també han estat desenvolupats en aquest llenguatge de programació.

La interfície d'usuari de l'aplicació ha estat desenvolupada per a *Android*, ja que la majoria de dispositius mòbils i tauletes utilitzen aquest sistema operatiu, així, la lògica implementada en l'aplicació està escrita en *Java*.

Finalment, els arxius de configuració i les funcions que ha de realitzar el servidor estan codificades en arxius *PHP*.

3 PLANIFICACIÓ INICIAL

3.1 Tasques

A l'iniciar el projecte es van definir una sèrie de tasques per dur a terme durant el seu desenvolupament. Aquesta definició pretenia establir una guia per a poder seguir els avenços que es produïen i comprovar que el projecte es podia desenvolupar en el temps establert. Cada tasca tenia associada un temps esperat per a ser completada en funció de la dificultat que podia suposar. Aquestes són les nou tasques en les que es va dividir el projecte, seguides d'una explicació sobre el que comprèn cadascuna:

1. Inici del treball, cerca d'informació i primeres reunions. Començar a definir el projecte, es va celebrar la primera reunió amb el tutor i es van començar a decidir aspectes rellevants (llenguatges de programació, estructura, etc.).
2. Codificació de funcions per identificar text en imatges. Cal destacar que aquestes funcions no han estat realitzades en aquest treball i que les va realitzar un altre estudiant [4]. Això és així per-

què si s'haguessin hagut d'implementar en aquest treball la complexitat del mateix hauria augmentat molt i no hauria sigut possible acabar-lo dins del termini previst. En aquesta tasca, doncs, es van interpretar les funcions donades i adaptar per a satisfer les necessitats del projecte. Aquestes funcions indicaven on es trobava el text en una imatge, a quines coordenades i ho ressaltaven amb un rectangle.

3. Codificació de funcions per a relacionar el text introduït per l'usuari amb la base de dades d'imatges. Com en la tasca anterior, s'han utilitzat funcions implementades per un altre estudiant [3] per a transformar la informació obtinguda en els rectangles de la tasca anterior en un text. Aquest text és el que es compara amb el text introduït per l'usuari.
4. Preparació del servidor i integració de les funcions per identificar text en imatges al servidor. Cal establir un servidor al que atacarem en l'aplicació, aquest servidor ha de ser accessible i integrar les funcions descrites anteriorment. Aquest servidor ha de contenir una base de dades que guardi la informació de totes les imatges allotjades al servidor.
5. Proves de comunicació amb el servidor. Verificar que el servidor és accessible i que rep les comunicacions enviades des de l'aplicació. Aquesta comunicació ha de ser ràpida i no s'ha d'enviar informació innecessària.
6. Disseny de la interfície. Una de les parts importants de l'aplicació és la interfície, caldrà pensar en un bon disseny i que sigui clar i entenedor. En cap cas ha de dificultar l'experiència de l'usuari.
7. Implementació de la interfície. Una vegada el disseny estigui validat caldrà implementar-lo i comprovar que s'adequa a les necessitats del projecte.
8. Proves finals. Amb totes les parts acabades caldrà verificar que funcionen correctament entre elles i que no hi ha errors.
9. Validació del producte final. Validar que el producte final és l'esperat, que compleix els objectius proposats i acabar de polir els últims detalls.

3.2 Recursos

En el desenvolupament del projecte s'han disposat dels següents recursos:

- L'estudiant que ha realitzat el projecte.
- Ordinador personal.
- Servidor per guardar les imatges.
- Entorn de programació Microsoft Visual Studio 2010 i 2013.
- Generador de documents Microsoft Office Word 2010.
- Eina per a la planificació Microsoft Project 2010.

3.3 Diagrama de Gantt

	Modo de	Número de tareas	Duración	Comienzo	Fin	Predecesoras
1		Inici del treball, cerca d'informació i primeres reunions.	2 sem.?	lun 10/02/14	vie 21/02/14	
2		Codificació de funcions per identificar text en imatges.	1 sem.?	lun 24/02/14	vie 28/02/14	1
3		Codificació de funcions per a relacionar el text introduït per l'usuari amb la base de dades d'imatges.	3 sem.?	lun 03/03/14	vie 21/03/14	1;2
4		Preparació del servidor i integració de les funcions per identificar text en imatges al servidor.	1 sem.?	lun 24/03/14	vie 28/03/14	3
5		Proves de comunicació amb el servidor.	1 sem.?	lun 31/03/14	vie 04/04/14	4
6		Disseny de la interfície.	1 sem.?	lun 07/04/14	vie 11/04/14	1
7		Implementació de la interfície.	2 sem.?	lun 14/04/14	vie 25/04/14	6
8		Proves finals.	3 sem.?	lun 28/04/14	vie 16/05/14	5;7
9		Validació del producte final.	2 sem.?	lun 19/05/14	vie 30/05/14	8

Fig. 1. Noms de les tasques, duració estimada, dates aproximades d'inici i final i dependències.

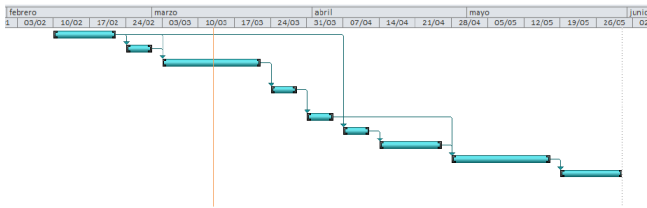


Fig. 2. Diagrama de Gantt de les tasques descrites en la fig. 1.

4 DESENVOLUPAMENT

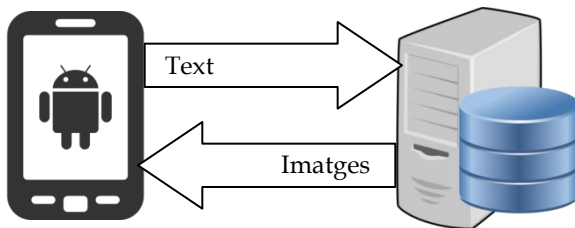


Fig. 3. Diagrama senzill del funcionament del projecte. L'aplicació envia un text al servidor i aquest retorna les cinc imatges.

4.1 Servidor

Com ja s'ha comentat, és necessari un servidor al que atacarem des de l'aplicació. Aquest servidor conté totes les imatges i una base de dades. En aquesta base de dades estan guardats tots els vectors (s'explicarà més endavant) amb el text que conté una imatge i a quina imatge i a quin rectangle d'aquella imatge pertany aquell text. Pot ser que una imatge tingui més d'un rectangle de text, bé, de fet és molt probable però això no ens interessa. És a dir, si trobem que un dels rectangles de la imatge coincideix amb el text enviat per l'usuari retornarem aquella imatge sense mirar si conté altres rectangles. Una imatge pot tenir molts rectangles de text però només ens n'interessarà un. A part de les imatges i la base de dades el servidor conté alguns arxius escrits en PHP. La majoria d'aquests arxius han servit per a fer proves o per a simplificar alguns arxius massa extensos, per això només en destacarem dos. El primer arxiu a destacar és un arxiu de configuració. Conté les dades per accedir a la base de dades, així com el nom de la base de dades, l'usuari i la contrasenya. Aquest arxiu és utilitzat pels altres a l'hora d'accedir a la base de dades (així el nom d'usuari i la contrasenya no estan repartits per tots els arxius, més seguretat i més comoditat en cas de voler canviar algun paràmetre). El segon arxiu a destacar és el codi encarregat de rebre el text entrat per l'usuari (des del codi de l'aplicació mòbil s'especifica a

quin fitxer s'envia el text). Aquest codi accedeix a la base de dades i recupera tots els vectors guardats, compara cadascun amb el text introduït i troba els cinc vectors que s'assemblen més. Quan tenim aquests cinc vectors trobem a quina imatge pertanyen i enviem les cinc imatges a l'aplicació. El servidor allotjarà les funcions per al tractament d'imatges, descrites en els subapartats següents.

4.1.1 Funcions per identificar text en imatges

Les funcions per identificar text en imatges troben rectangles que poden contenir text en les imatges que analitzen i retornen la localització d'aquests rectangles. Com ja s'ha introduït en l'explicació de les tasques realitzades, aquestes funcions han estat implementades per un altre estudiant i estan explicades a [4].

Aquestes funcions no s'han pogut implementar per dos motius. El primer és que la identificació del text que apareix en imatges pertany a la menció de computació i aquest treball pertany a la menció d'enginyeria del software. Aquest treball està més centrat en la creació i la gestió de l'aplicació. El segon motiu és que si s'hagués optat per implementar aquestes funcions de zero molt probablement no s'hauria aconseguit i el treball restaria incomplet.

Tot i tenir aquestes funcions ja implementades, adaptar-les al projecte va ser una de les tasques que més problemes va comportar. L'aparició de molts errors i la dificultat de corregir-los va suposar un greu endarreriment en la planificació inicial i va implicar establir una nova planificació.

Convé destacar que la detecció de text no és del tot exacte. El text en qüestió pot estar lleugerament borrós, pot estar parcialment tapat per algun altre objecte en la imatge, pot tenir diferents mides, estils de lletra, etc. Per això direm que aquestes funcions proporcionen una aproximació no correcta al 100% però sí molt pròxima.

4.1.2 Càlcul de vectors per cada rectangle

Amb els rectangles obtinguts de les funcions de l'apartat anterior es calculen vectors que contenen les probabilitats de que apareguin unes lletres o unes altres. D'aquesta manera s'intenta preveure quines paraules poden contenir aquests rectangles. A la base de dades guardarem cada vector juntament amb el rectangle associat i la imatge a la que pertany. Quan trobem un vector que és semblant al text enviat necessitem saber a quina imatge pertany. Per motius semblants als anteriors descrits, aquestes funcions no han estat realitzades en aquest treball sinó que han estat realitzades per un altre estudiant a [3]. pròxima.

4.1.3 Cerca

Al rebre el text enviat per l'usuari, el servidor executarà funcions semblants a les del punt anterior per obtenir un vector de probabilitats. Amb una consulta a la base de dades es recuperaran tots els vectors guardats i es multiplicaran un per un amb el vector obtingut del text. Com més alt sigui el valor obtingut més semblant seran els dos vectors (cerca per semblança). Finalment només caldrà buscar les cinc imatges que corresponguin als cinc vectors més semblants i enviar-les a l'aplicació.

4.2 Aplicació

Per a no sobrecarregar el dispositiu amb masses càlculs aquests es faran al servidor. Així, la funcionalitat de l'aplicació serà el més senzill possible. Quan l'usuari iniciï l'aplicació veurà que la pantalla inicial ja és la pantalla per introduir text (no és necessari utilitzar pantalles de benvinguda ni semblants ja que destorbarien més que no pas ajudarien). Un missatge en forma de títol indicarà a l'usuari que introdueixi el text que vol buscar. Aquesta pantalla disposa del camp per introduir-lo i de dos botons, un per a buscar i l'altre per a sortir (el primer envia les dades i el segon finalitza l'execució de l'aplicació). Al prémer el botó de buscar l'aplicació espera rebre les cinc imatges. Quan les ha rebut es canvia de pantalla. Aquesta segona pantalla està dividida en dues seccions. A la part esquerra l'usuari pot veure una imatge en gran. A la part dreta l'usuari veu les cinc imatges i un botó de tornar enrere. Si l'usuari prem qualsevol de les cinc imatges aquesta serà la que es veurà en gran a la part esquerra. Si prem el botó de tornar enrere serà portat a la primera pantalla.

4.2.1 Interfície

Explicada la funcionalitat de l'aplicació el següent punt és mostrar captures de pantalla per a poder veure la interfície. Un dels objectius del projecte era aconseguir una interfície clara i entenedora, que permetés a l'usuari utilitzar l'aplicació sense problemes i de forma intuïtiva. La senzillesa de la interfície desenvolupada satisfà aquest objectiu.

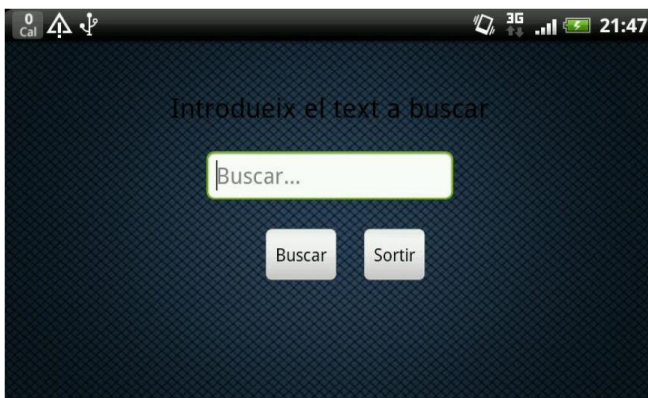


Fig. 4. Aquest és l'aspecte de la primera pantalla. El títol s'usa de missatge per indicar a l'usuari el que pot fer. Disposa d'un camp per introduir text i els dos botons comentats, el botó buscar que envia el text introduït i el botó sortir que serveix per tancar l'aplicació.

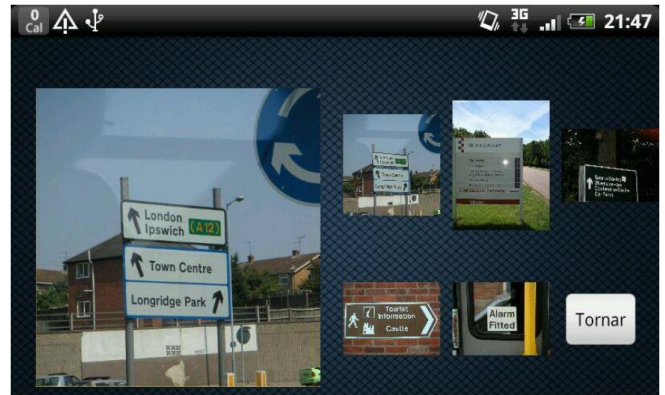


Fig. 5. Aquí es mostra la segona pantalla. Les dues seccions no estan delimitades però s'aprecien fàcilment.

Aquesta segona pantalla té alguns aspectes a destacar. Quan l'usuari premi qualsevol de les imatges aquesta es mostrarà en gran a la part esquerra però, al fer el canvi de pantalla, com que l'usuari no ha premut cap imatge la que es mostrarà en gran serà la primera imatge rebuda. La raó d'això és que la primera imatge rebuda serà la que tingui més coincidències amb el que l'usuari estava buscant.

Un altre aspecte important és que la interfície està forçada per a ser vista en horitzontal (*landscape* en *Android*). La primera idea era que fos vertical, però dificultava mostrar correctament les imatges. Si les miniatures de les imatges estaven en una sola línia gairebé no s'apreciaven (en dispositius amb pantalles més petites directament no s'haurien pogut apreciar) i si estaven en dues línies restaven massa espai a la imatge ampliada. Una de les propostes per a solucionar aquest problema va ser utilitzar un tema horitzontal i forçar que no es podés tornar vertical. A l'implementar aquesta proposta vam veure que teníem la interfície que volíem (mostrava les imatges correctament, es podien apreciar les dues mides sense problemes i satisfia l'objectiu de clara i entenedora).

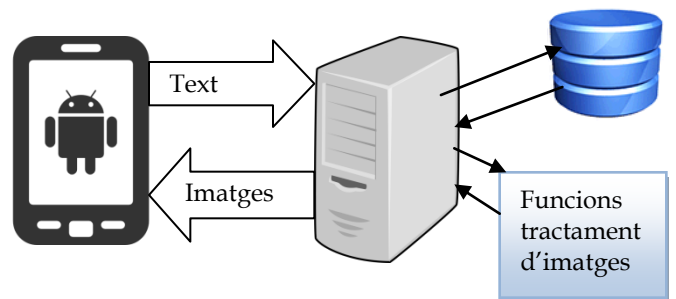


Fig. 6. Diagrama complet del funcionament del projecte. L'aplicació envia un text al servidor. El servidor fa una consulta a la base de dades per obtenir tots els vectors. Després converteix el text en un vector i el compara amb els altres per obtenir els cinc vectors més semblants. Torna a consultar la base de dades i obté les imatges corresponents. Finalment envia les imatges a l'aplicació.

5 RESULTATS

5.1 Canvis en la planificació

Van aparèixer problemes en les tasques 2 i 3 que van tardar a solucionar-se. Això va obligar a allargar enormement la seva duració. Aquests problemes van provocar l'endarreriment del projecte perquè la majoria de tasques següents depenien de la consecució d'aquestes dues. Sense les funcions que s'havien de desenvolupar en aquelles tasques no es podia posar en marxa el servidor ni fer les proves de connexió perquè aquelles funcions eren la lògica que havia d'implementar el servidor. Allargar aquestes tasques també va afectar a la data final prevista, però seguia dins del període acadèmic previst. Això va ser possible perquè inicialment, a l'establir la planificació, es va acordar una data d'entrega amb suficient marge respecte a la data d'entrega final per si hi havia algun contratemps com aquest.

Modo de	Nombre de tarea	Duración	Comienzo	Fin	Predecesoras
1	Inici del treball, cerca d'informació i primeres reunions.	2 sem.	lun 10/02/14	vie 21/02/14	
2	Codificació de funcions per identificar text en imatges.	8 sem.	lun 24/02/14	vie 18/04/14	1
3	Codificació de funcions per a relacionar el text introduït per l'usuari amb la base de dades d'imatges.	8 sem.	lun 03/03/14	vie 25/04/14	1
4	Preparació del servidor i integració de les funcions per identificar text en imatges al servidor.	1 sem?	lun 28/04/14	vie 02/05/14	3;2
5	Proves de comunicació amb el servidor.	1 sem?	lun 05/05/14	vie 09/05/14	4
6	Disseny de la interfície.	1 sem?	lun 07/04/14	vie 11/04/14	1
7	Implementació de la interfície.	2 sem?	lun 14/04/14	vie 25/04/14	6
8	Proves finals.	3 sem.	lun 12/05/14	vie 30/05/14	5;7
9	Validació del producte final.	2 sem.	lun 02/06/14	vie 13/06/14	8

Fig. 7. Reestructuració de les dates de les tasques.

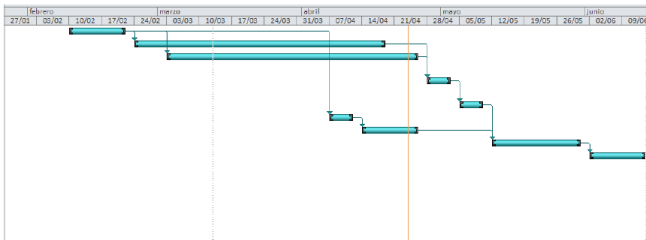


Fig. 8. Nou diagrama de Gantt després de la reestructuració.

5.2 Problemes amb els llenguatges

Des de l'aplicació (codi *Android* i Java) s'envia un text a un arxiu allotjat al servidor (codi PHP) que crida a les funcions d'un arxiu (codi C++ fet en Windows i executant-se sobre un servidor Linux) del mateix servidor. Aquesta execució sona molt enrevessada però no es pot simplificar, l'aplicació demana un codi *Android* i Java, el servidor rep les peticions en PHP i les funcions per calcular els vectors estan en C++. Això ha portat problemes com la impossibilitat de retornar una variable de tipus *float* des del codi C++ (per estàndard la funció *main* en C++ ha de retornar un valor *int*). Aquest problema redueix l'exactitud de les funcions però no afecta al concepte pensat, però el segon problema sí. El segon problema també té a veure amb el codi en C++ i és la impossibilitat de retornar un vector (amb retornar em torno a referir a la funció *int*) al finalitzar l'execució del programa i utilitzar-lo en un altre (en aquest cas utilitzar el vector obtingut del codi C++ en el codi PHP). No s'ha trobat solució al problema i s'ha optat per un canvi de concepte: en lloc de guardar els vectors a la base de dades es guardaran cadenes de text per cada imatge.

Des del servidor, al rebre un text de l'usuari, es fa una consulta a la base de dades i es recuperen totes les cadenes de text (i no els vectors com es tenia previst originalment). Aquestes cadenes s'envien una a una a l'executable obtingut de compilar el codi en C++ (que com ja s'ha comentat ha estat implementat per un altre estudiant) juntament amb el text de l'usuari. Aquest codi obté els vectors de les cadenes i en calcula el producte escalar. Aquest nombre és un nombre real, però degut a les limitacions del llenguatge C++ s'ha de retornar com un nombre enter (aquesta pèrdua d'informació redueix l'exactitud de la cerca).

Tornant al codi del servidor, a mesura que es van rebent aquests productes es van guardant els cinc més grans junt amb la imatge a la que pertany la cadena comparada. Quan s'han enviat i rebut totes es retorna la informació sobre les cinc imatges més semblants a l'aplicació per a que les pugui mostrar a l'usuari.

Aquest canvi de concepte ha fet que el projecte no fos tant automàtic, ja que es necessita que algú llegeixi les imatges per a obtenir la cadena de text. Però no és una tasca que recau únicament sobre el responsable de les imatges, ja que primer s'analitzen amb el codi obtingut a [4] i només es procedeix a identificar les zones de text que es retornen. Si alguna zona de text no ha estat identificada per aquestes funcions no entrarà a la base de dades.

5.3 Exemples d'execució

Com ja s'ha comentat durant l'article la cerca d'imatges no és exacta. Si a això hi sumem que la base de dades utilitzada conté deu imatges (amb un total de setanta-tres cadenes de text) es pot donar el cas que el text enviat per l'usuari no correspongui amb cap cadena de text. En aquest cas les cinc imatges més semblants no tindran res a veure amb el text enviat.

Després de múltiples proves per comprovar el funcionament de l'aplicació es destaquen tres exemples:

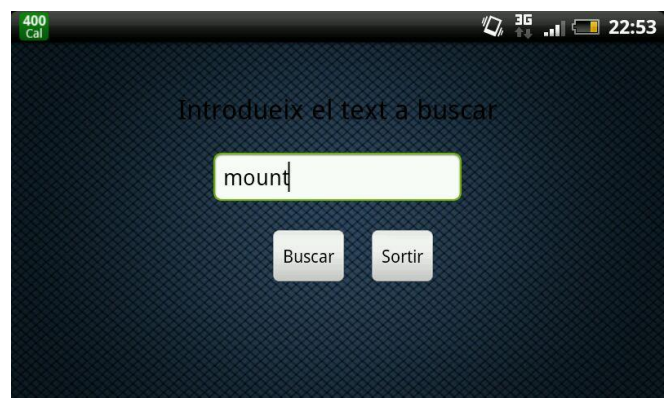


Fig. 9. S'introdueix "mount" com a text a buscar. Les tres cerques estan preparades i s'han fet amb cadenes que estan a la base de dades, sinó no es podrien comprovar.

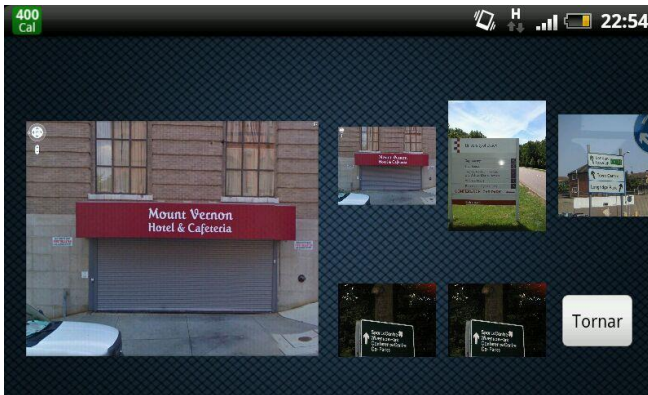


Fig. 10. Resultat de la cerca. La primera imatge és la que esperàvem rebre.

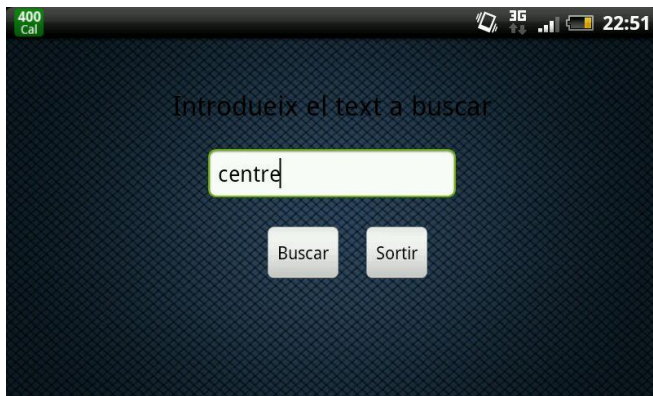


Fig. 11. Segon exemple, s'introdueix "centre" com a text a buscar.

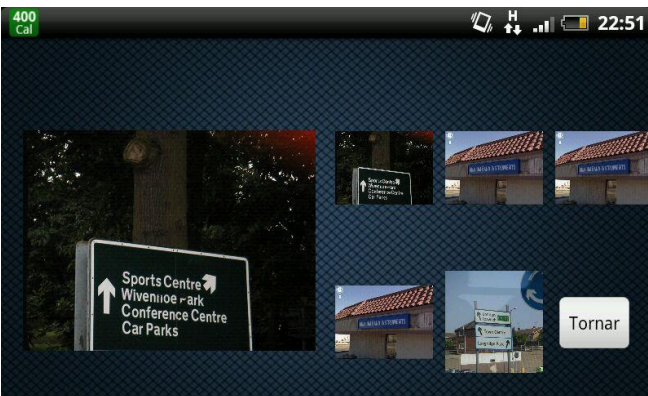


Fig. 12. Resultat de la segona cerca. En aquest cas la primera imatge també és l'esperada.

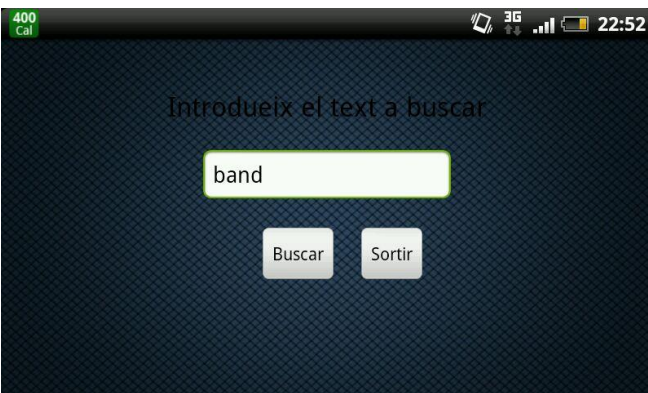


Fig. 13. En la tercera cerca es buscarà "band".

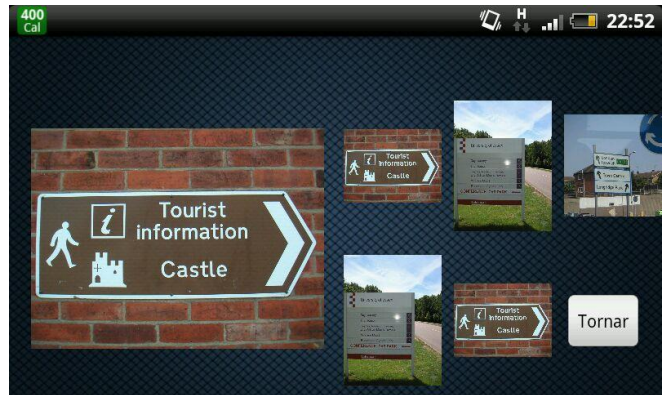


Fig. 14. Aquest és el resultat de la tercera cerca. Tot i tenir una cadena "band" a la base de dades no ens l'ha retornat. Aquesta execució és un dels exemples que demostren que la cerca no sempre és fiable.

Ja per acabar amb els resultats, destacar que es podria "jugar" a buscar cadenes de text que no estiguin a la base de dades per veure què interpreta i què ens retorna l'aplicació. Aquests casos, però, no es podran tenir en compte a l'hora d'avaluar la veracitat dels resultats.

6 CONCLUSIÓ

A l'acabar el projecte cal donar un cop d'ull als objectius plantejats i veure si s'han assolit. Tot i les múltiples dificultats que han anat sorgint els tres grans objectius s'han complert, l'aplicació i el servidor són una realitat i la comunicació entre ells és eficient i fiable. Com a contres es podria destacar el canvi de concepte explicat en la secció 5.2. És un punt negatiu perquè ha comportat el replanteig d'alguns conceptes però no ha evitat que es complissin els objectius.

Mirant al futur, el projecte es podria continuar de moltes maneres. Es podria optar per millorar l'eficiència de les cerques per a fer més fiable l'aplicació o donar-li més funcions a l'usuari. Possibles exemples de funcions podrien ser la possibilitat d'afegir imatges a la base de dades des de l'aplicació o el poder triar quantes imatges ens retorna l'aplicació. Una altra continuació podria ser implementar la mateixa funcionalitat però a través d'una web (que com s'ha discutit a l'inici de l'article és una possibilitat que es va descartar per falta de temps).

AGRAÏMENTS

Un projecte tant important com aquest no es podria tancar sense un petit reconeixement a les persones que m'han ajudat a tirar endavant i fer-lo possible.

Primer de tot vull agrair a l'Ernest Valveny que hagi sigut el meu tutor i que m'hagi guiat i ajudat en tot moment, que s'hagi reunit amb mi i m'hagi ajudat a entendre conceptes del projecte que eren nous per a mi, que m'hagi aguantat i m'hagi explicat el què volia i on m'equivocava. El projecte va sorgir d'una idea seva i espero haver-la satisfet. Vull agrair al Lluís Gómez que hagi implementat les funcions per a identificar zones de text en les imatges. Vull agrair al Jon Almazán que s'hagi encarregat del codi per obtenir els vectors a partir de les cadenes de text, de totes les funcions que ha implementat i dels dubtes que m'ha aclarit. Sense vosaltres tres no hauria sigut capaç d'acabar el projecte, gràcies de veritat.

No voldria tancar aquesta secció sense agrair a la meua família i amics la paciència que han tingut amb mi, que m'hagin donat suport en tot moment i que hagin cregut en mi. Gràcies a vosaltres he estat capaç de tirar endavant.

BIBLIOGRAFIA

- [1] OPENCV DEV TEAM. (2014, Mar). *Scene Text Detection* [Online]. Available: <http://docs.opencv.org/trunk/modules/objdetect/doc/erfilter.html>
- [2] NEUMANN L., MATAS J. *Real-Time Scene Text Localization and Recognition*. CVPR, 2012.
- [3] ALMAZAN J., GORDO A., FORNES A., VALVENY E. *Handwritten Word Spotting with Corrected Attributes*. 15th IEEE International Conference on Computer Vision, 2013.
- [4] GÓMEZ LL., KARATZAS D. *Multi-script Text Extraction from Natural Scenes* (p. 467-471). 12th International Conference on Document Analysis and Recognition, 2013.
- [5] Referència 5

Estàndards, normatives i reglaments

ISO 9001: Estàndard que descriu el sistema de qualitat utilitzat per mantenir el desenvolupament d'un producte que impliqui disseny (en el cas d'aquest projecte, disseny d'un producte software).